CHAPTER 7

7.1. Write (with probability approaching one)

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right).$$

From SOLS.2, the weak law of large numbers, and Slutsky's Theorem,

$$\text{plim}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1} = \mathbf{A}^{-1}.$$

Further, under SOLS.1, the WLLN implies that plim $\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right) = \mathbf{0}$. Thus,

$$\text{plim}\ \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \text{plim}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\cdot\text{plim}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right) = \boldsymbol{\beta} + \mathbf{A}^{-1}\cdot\mathbf{0} = \boldsymbol{\beta}. \quad \blacksquare$$

7.3. a. Since OLS equation-by-equation is the same as GLS when $\boldsymbol{\Omega}$ is diagonal, it suffices to show that the GLS estimators for different equations are asymptotically uncorrelated. This follows if the asymptotic variance matrix is block diagonal (see Section 3.5), where the blocking is by the parameter vector for each equation. To establish block diagonality, we use the result from Theorem 7.4: under SGLS.1, SGLS.2, and SGLS.3,

$$\text{Avar}\ \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = [\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)]^{-1}.$$

Now, we can use the special form of $\mathbf{X}_i$ for SUR (see Example 7.1), the fact that $\boldsymbol{\Omega}^{-1}$ is diagonal, and SGLS.3. In the SUR model with diagonal $\boldsymbol{\Omega}$, SGLS.3 implies that $\text{E}(u_{ig}^2\mathbf{x}_{ig}'\mathbf{x}_{ig}) = \sigma_g^2\text{E}(\mathbf{x}_{ig}'\mathbf{x}_{ig})$ for all $g = 1,\ldots,G$, and $\text{E}(u_{ig}u_{ih}\mathbf{x}_{ig}'\mathbf{x}_{ih}) = \text{E}(u_{ig}u_{ih})\text{E}(\mathbf{x}_{ig}'\mathbf{x}_{ih}) = \mathbf{0}$, all $g \neq h$. Therefore, we have

$$\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = \begin{pmatrix} \sigma_1^{-2}\text{E}(\mathbf{x}_{i1}'\mathbf{x}_{i1}) & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \cdot & & \\ & & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \sigma_G^{-2}\text{E}(\mathbf{x}_{iG}'\mathbf{x}_{iG}) \end{pmatrix}.$$

When this matrix is inverted, it is also block diagonal. This shows that the asymptotic variance of what we wanted to show.

b. To test any linear hypothesis, we can either construct the Wald statistic or we can use the weighted sum of squared residuals form of the statistic as in (7.52) or (7.53). For the restricted SSR we must estimate the model with the restriction $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ imposed. See Problem 7.6 for one way to impose general linear restrictions.

c. When $\boldsymbol{\Omega}$ is diagonal in a SUR system, system OLS and GLS are the same. Under SGLS.1 and SGLS.2, GLS and FGLS are asymptotically equivalent (regardless of the structure of $\boldsymbol{\Omega}$) whether or not SGLS.3 holds. But, if $\hat{\boldsymbol{\beta}}_{SOLS} = \hat{\boldsymbol{\beta}}_{GLS}$ and $\sqrt{N}(\hat{\boldsymbol{\beta}}_{FGLS} - \hat{\boldsymbol{\beta}}_{GLS}) = o_p(1)$, then $\sqrt{N}(\hat{\boldsymbol{\beta}}_{SOLS} - \hat{\boldsymbol{\beta}}_{FGLS}) = o_p(1)$. Thus, when $\boldsymbol{\Omega}$ is diagonal, OLS and FGLS are asymptotically equivalent, even if $\hat{\boldsymbol{\Omega}}$ is estimated in an unrestricted fashion and even if the system homoskedasticity assumption SGLS.3 does not hold.

7.5. This is easy with the hint. Note that

$$\left(\hat{\boldsymbol{\Omega}}^{-1} \otimes \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)\right)^{-1} = \hat{\boldsymbol{\Omega}} \otimes \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}.$$

Therefore,

$$\hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\Omega}} \otimes \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}\right)(\hat{\boldsymbol{\Omega}}^{-1} \otimes \mathbf{I}_K)\begin{pmatrix} \sum_{i=1}^{N} \mathbf{x}_i' y_{i1} \\ \vdots \\ \sum_{i=1}^{N} \mathbf{x}_i' y_{iG} \end{pmatrix} = \left(\mathbf{I}_G \otimes \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}\right)\begin{pmatrix} \sum_{i=1}^{N} \mathbf{x}_i' y_{i1} \\ \vdots \\ \sum_{i=1}^{N} \mathbf{x}_i' y_{iG} \end{pmatrix}.$$

Straightforward multiplication shows that the right hand side of the equation is just the vector of stacked $\hat{\hat{\boldsymbol{\beta}}}_g$, $g = 1,\ldots,G$. where $\hat{\hat{\boldsymbol{\beta}}}_g$ is the OLS estimator for equation g. ■

7.7. a. First, the diagonal elements of $\boldsymbol{\Omega}$ are easily found since $E(u_{it}^2) = E[E(u_{it}^2|\mathbf{x}_{it})] = \sigma_t^2$ by iterated expectations. Now, consider $E(u_{it}u_{is})$, and take $s < t$ without loss of generality. Under (7.79), $\mathrm{E}(u_{it}|u_{is}) = 0$ since u_{is} is a subset of the conditioning information in (7.80). Applying the law of iterated expectations (LIE) again we have $\mathrm{E}(u_{it}u_{is}) = \mathrm{E}[\mathrm{E}(u_{it}u_{is}|u_{is})] = \mathrm{E}[\mathrm{E}(u_{it}|u_{is})u_{is})] = 0$.

b. The GLS estimator is

$$\boldsymbol{\beta}^{*} \equiv \left(\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{y}_i\right)$$
$$= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_t^{-2}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_t^{-2}\mathbf{x}_{it}'y_{it}\right).$$

c. If, say, $y_{it} = \beta_0 + \beta_1 y_{i,t-1} + u_{it}$, then y_{it} is clearly correlated with u_{it}, which says that $\mathbf{x}_{i,t+1} = y_{it}$ is correlated with u_{it}. Thus, SGLS.1 does not hold. Generally, SGLS.1 holds whenever there is feedback from y_{it} to $\mathbf{x}_{is}$, $s > t$. However, since $\boldsymbol{\Omega}^{-1}$ is diagonal, $\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i = \sum_{t=1}^{T}\mathbf{x}_{it}'\sigma_t^{-2}u_{it}$, and so

$$\mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i) = \sum_{t=1}^{T}\sigma_t^{-2}\mathrm{E}(\mathbf{x}_{it}'u_{it}) = \mathbf{0}$$

since $\mathrm{E}(\mathbf{x}_{it}'u_{it}) = \mathbf{0}$ under (7.80). Thus, GLS is consistent in this case without SGLS.1.

d. First, since $\boldsymbol{\Omega}^{-1}$ is diagonal, $\mathbf{X}_i'\boldsymbol{\Omega}^{-1} = (\sigma_1^{-2}\mathbf{x}_{i1}', \sigma_2^{-2}\mathbf{x}_{i2}', \ldots, \sigma_T^{-2}\mathbf{x}_{iT}')'$, and so

$$\mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = \sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_t^{-2}\sigma_s^{-2}\mathrm{E}(u_{it}u_{is}\mathbf{x}_{it}'\mathbf{x}_{is}).$$

First consider the terms for $s \neq t$. Under (7.80), if $s < t$, $\mathrm{E}(u_{it}|\mathbf{x}_{it},u_{is},\mathbf{x}_{is}) = 0$, and so by the LIE, $\mathrm{E}(u_{it}u_{is}\mathbf{x}_{it}'\mathbf{x}_{is}) = \mathbf{0}$, $t \neq s$. Next, for each t,

$$\mathrm{E}(u_{it}^2\mathbf{x}_{it}'\mathbf{x}_{it}) = \mathrm{E}[\mathrm{E}(u_{it}^2\mathbf{x}_{it}'\mathbf{x}_{it}|\mathbf{x}_{it})] = \mathrm{E}[\mathrm{E}(u_{it}^2|\mathbf{x}_{it})\mathbf{x}_{it}'\mathbf{x}_{it})]$$
$$= \mathrm{E}[\sigma_t^2\mathbf{x}_{it}'\mathbf{x}_{it}] = \sigma_t^2\mathrm{E}(\mathbf{x}_{it}'\mathbf{x}_{it}), \quad t = 1,2,\ldots,T.$$

It follows that

$$\mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = \sum_{t=1}^{T}\sigma_t^{-2}\mathrm{E}(\mathbf{x}_{it}'\mathbf{x}_{it}) = \mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i).$$

e. First, run pooled regression across all i and t; let $\hat{\hat{u}}_{it}$ denote the pooled OLS residuals. Then, for each t, define

$$\hat{\sigma}_t^2 = N^{-1}\sum_{i=1}^{N}\hat{\hat{u}}_{it}^2$$

(We might replace N with $N - K$ as a degrees-of-freedom adjustment.) Then, by standard arguments, $\hat{\sigma}_t^2 \xrightarrow{p} \sigma_t^2$ as $N \to \infty$.

f. We have verified the assumptions under which standard FGLS statistics have nice properties (although we relaxed SGLS.1). In particular, standard errors obtained from (7.51) are asymptotically valid, and F statistics from (7.53) are valid. Now, if $\hat{\mathbf{\Omega}}$ is taken to be the diagonal matrix with $\hat{\sigma}_t^2$ as the t^{th} diagonal, then the FGLS statistics are easily shown to be identical to the statistics obtained by performing pooled OLS on the equation

$$(y_{it}/\hat{\sigma}_t) = (\mathbf{x}_{it}/\hat{\sigma}_t)\boldsymbol{\beta} + error_{it},\ t = 1,2,\ldots,T,\ i = 1,\ldots,N.$$

We can obtain valid standard errors, t statistics, and F statistics from this weighted least squares analysis. For F testing, note that the $\hat{\sigma}_t^2$ should be obtained from the pooled OLS residuals for the unrestricted model.

g. If $\sigma_t^2 = \sigma^2$ for all $t = 1,\ldots,T$, inference is very easy. FGLS reduces to pooled OLS. Thus, we can use the standard errors and test statistics reported by a standard OLS regression pooled across i and t.

7.9. The Stata session follows. I first test for serial correlation before computing the fully robust standard errors:

```
. reg lscrap d89 grant grant_1 lscrap_1 if year != 1987

  Source |       SS       df       MS                  Number of obs =     108
---------+------------------------------               F(  4,   103) =  153.67
   Model |  186.376973     4  46.5942432               Prob > F      =  0.0000
Residual |  31.2296502   103  .303200488               R-squared     =  0.8565
---------+------------------------------               Adj R-squared =  0.8509
   Total |  217.606623   107  2.03370676               Root MSE      =  .55064
```

```
------------------------------------------------------------------------------
  lscrap |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     d89 |  -.1153893   .1199127     -0.962   0.338      -.3532078    .1224292
   grant |  -.1723924   .1257443     -1.371   0.173      -.4217765    .0769918
 grant_1 |  -.1073226   .1610378     -0.666   0.507       -.426703    .2120579
lscrap_1 |   .8808216   .0357963     24.606   0.000        .809828    .9518152
   _cons |  -.0371354   .0883283     -0.420   0.675      -.2123137     .138043
------------------------------------------------------------------------------
```

The estimated effect of *grant*, and its lag, are now the expected sign, but neither is strongly statistically significant. The variable *grant* would be if we use a 10% significance level and a one-sided test. The results are certainly different from when we omit the lag of log(*scrap*).

Now test for AR(1) serial correlation:

```
. predict uhat, resid
(363 missing values generated)

. gen uhat_1 = uhat[_n-1] if d89
(417 missing values generated)

. reg lscrap grant grant_1 lscrap_1 uhat_1 if d89

  Source |       SS       df       MS                  Number of obs =      54
---------+------------------------------               F(  4,    49) =   73.47
   Model |  94.4746525     4  23.6186631               Prob > F      =  0.0000
Residual |  15.7530202    49  .321490208               R-squared     =  0.8571
---------+------------------------------               Adj R-squared =  0.8454
   Total |  110.227673    53  2.07976741               Root MSE      =    .567

------------------------------------------------------------------------------
  lscrap |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   grant |   .0165089    .215732      0.077   0.939      -.4170208    .4500385
 grant_1 |  -.0276544   .1746251     -0.158   0.875      -.3785767    .3232679
lscrap_1 |   .9204706   .0571831     16.097   0.000       .8055569    1.035384
  uhat_1 |   .2790328   .1576739      1.770   0.083      -.0378247    .5958904
   _cons |   -.232525   .1146314     -2.028   0.048      -.4628854   -.0021646
------------------------------------------------------------------------------

. reg lscrap d89 grant grant_1 lscrap_1 if year != 1987, robust cluster(fcode)

Regression with robust standard errors                 Number of obs =     108
                                                       F(  4,    53) =   77.24
                                                       Prob > F      =  0.0000
```

```
                                                       R-squared     =  0.8565
Number of clusters (fcode) = 54                        Root MSE      =  .55064

------------------------------------------------------------------------------
             |               Robust
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d89 |  -.1153893   .1145118    -1.01   0.318    -.3450708    .1142922
       grant |  -.1723924   .1188807    -1.45   0.153    -.4108369    .0660522
     grant_1 |  -.1073226   .1790052    -0.60   0.551    -.4663616    .2517165
    lscrap_1 |   .8808216   .0645344    13.65   0.000     .7513821    1.010261
       _cons |  -.0371354   .0893147    -0.42   0.679     -.216278    .1420073
------------------------------------------------------------------------------
```

The robust standard errors for *grant* and $grant_{-1}$ are actually smaller than the usual ones, making both more statistically significant. However, *grant* and $grant_{-1}$ are jointly insignificant:

```
. test grant grant_1

 ( 1)  grant = 0.0
 ( 2)  grant_1 = 0.0

       F(  2,    53) =    1.14
            Prob > F =    0.3266
```

7.11. a. The following Stata output should be self-explanatory. There is strong evidence of positive serial correlation in the static model, and the fully robust standard errors are much larger than the nonrobust ones.

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d82-d87

  Source |       SS       df       MS                  Number of obs =     630
---------+------------------------------               F( 11,   618) =   74.49
   Model |  117.644669    11  10.6949699               Prob > F      =  0.0000
Residual |   88.735673   618  .143585231               R-squared     =  0.5700
---------+------------------------------               Adj R-squared =  0.5624
   Total |  206.380342   629  .328108652               Root MSE      =  .37893

------------------------------------------------------------------------------
 lcrmrte |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 lprbarr |  -.7195033   .0367657    -19.570   0.000      -.7917042   -.6473024
lprbconv |  -.5456589   .0263683    -20.694   0.000      -.5974413   -.4938765
lprbpris |   .2475521   .0672268      3.682   0.000       .1155314    .3795728
```

```
 lavgsen |  -.0867575   .0579205     -1.498   0.135      -.2005023    .0269872
  lpolpc |   .3659886   .0300252     12.189   0.000       .3070248    .4249525
     d82 |   .0051371    .057931      0.089   0.929      -.1086284    .1189026
     d83 |   -.043503   .0576243     -0.755   0.451      -.1566662    .0696601
     d84 |  -.1087542    .057923     -1.878   0.061       -.222504    .0049957
     d85 |  -.0780454   .0583244     -1.338   0.181      -.1925835    .0364927
     d86 |  -.0420791   .0578218     -0.728   0.467        -.15563    .0714718
     d87 |  -.0270426    .056899     -0.475   0.635      -.1387815    .0846963
   _cons |  -2.082293   .2516253     -8.275   0.000      -2.576438   -1.588149
------------------------------------------------------------------------------

. predict uhat, resid

. gen uhat_1 = uhat[_n-1] if year > 81
(90 missing values generated)

. reg uhat uhat_1

  Source |       SS       df       MS                  Number of obs =     540
---------+------------------------------               F(  1,   538) =  831.46
   Model |  46.6680407     1  46.6680407               Prob > F      =  0.0000
Residual |  30.1968286   538  .056127934               R-squared     =  0.6071
---------+------------------------------               Adj R-squared =  0.6064
   Total |  76.8648693   539  .142606437               Root MSE      =  .23691

------------------------------------------------------------------------------
    uhat |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  uhat_1 |   .7918085     .02746     28.835   0.000       .7378666    .8457504
   _cons |   1.74e-10   .0101951      0.000   1.000      -.0200271    .0200271
------------------------------------------------------------------------------
```

Because of the strong serial correlation, I obtain the fully robust standard errors:

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d82-d87, robust
cluster(county)

Regression with robust standard errors                 Number of obs =     630
                                                       F( 11,    89) =   37.19
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5700
Number of clusters (county) = 90                       Root MSE      =  .37893

------------------------------------------------------------------------------
             |               Robust
     lcrmrte |      Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.7195033   .1095979    -6.56   0.000    -.9372719   -.5017347
```

```
    lprbconv |  -.5456589   .0704368    -7.75   0.000    -.6856152   -.4057025
    lprbpris |   .2475521   .1088453     2.27   0.025     .0312787    .4638255
     lavgsen |  -.0867575   .1130321    -0.77   0.445    -.3113499    .1378348
      lpolpc |   .3659886    .121078     3.02   0.003     .1254092    .6065681
         d82 |   .0051371   .0367296     0.14   0.889    -.0678438    .0781181
         d83 |   -.043503    .033643    -1.29   0.199    -.1103509    .0233448
         d84 |  -.1087542   .0391758    -2.78   0.007    -.1865956   -.0309127
         d85 |  -.0780454   .0385625    -2.02   0.046    -.1546683   -.0014224
         d86 |  -.0420791   .0428788    -0.98   0.329    -.1272783    .0431201
         d87 |  -.0270426   .0381447    -0.71   0.480    -.1028353    .0487502
       _cons |  -2.082293   .8647054    -2.41   0.018    -3.800445   -.3641423
------------------------------------------------------------------------------

. drop uhat uhat_1
```

b. We lose the first year, 1981, when we add the lag of log(*crmrte*):

```
. gen lcrmrt_1 = lcrmrte[_n-1] if year > 81
(90 missing values generated)

. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d83-d87 lcrmrt_1

  Source |       SS       df       MS                  Number of obs =     540
---------+------------------------------               F( 11,   528) =  464.68
   Model |  163.287174    11  14.8442885               Prob > F      =  0.0000
Residual |  16.8670945   528  .031945255               R-squared     =  0.9064
---------+------------------------------               Adj R-squared =  0.9044
   Total |  180.154268   539  .334237975               Root MSE      =  .17873

------------------------------------------------------------------------------
 lcrmrte |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 lprbarr |  -.1668349   .0229405     -7.273   0.000       -.2119007   -.1217691
lprbconv |  -.1285118   .0165096     -7.784   0.000       -.1609444   -.0960793
lprbpris |  -.0107492   .0345003     -0.312   0.755        -.078524    .0570255
 lavgsen |  -.1152298    .030387     -3.792   0.000        -.174924   -.0555355
  lpolpc |    .101492   .0164261      6.179   0.000        .0692234    .1337606
     d83 |  -.0649438   .0267299     -2.430   0.015       -.1174537   -.0124338
     d84 |  -.0536882   .0267623     -2.006   0.045       -.1062619   -.0011145
     d85 |  -.0085982   .0268172     -0.321   0.749       -.0612797    .0440833
     d86 |   .0420159    .026896      1.562   0.119       -.0108203    .0948522
     d87 |   .0671272   .0271816      2.470   0.014        .0137298    .1205245
lcrmrt_1 |   .8263047   .0190806     43.306   0.000        .7888214    .8637879
   _cons |  -.0304828   .1324195     -0.230   0.818       -.2906166     .229651
------------------------------------------------------------------------------
```

Not surprisingly, the lagged crime rate is very significant. Further, including it makes all other coefficients much smaller in magnitude. The variable log(*prbpris*) now has a negative sign, although it is insignificant. We still get a positive relationship between size of police force and crime rate, however.

c. There is no evidence of serial correlation in the model with a lagged dependent variable:

```
. predict uhat, resid
(90 missing values generated)

. gen uhat_1 = uhat[_n-1] if year > 82
(180 missing values generated)

. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d84-d87 lcrmrt_1 uhat_1
```

From this regression the coefficient on $uhat_{-1}$ is only −.059 with *t* statistic −.986, which means that there is little evidence of serial correlation (especially since $\hat{\rho}$ is practically small). Thus, I will not correct the standard errors.

d. None of the log(*wage*) variables is statistically significant, and the magnitudes are pretty small in all cases:

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d83-d87 lcrmrt_1 lwcon-
lwloc

  Source |       SS       df       MS                  Number of obs =     540
---------+------------------------------               F( 20,   519) =  255.32
   Model |  163.533423    20  8.17667116               Prob > F      =  0.0000
Residual |  16.6208452   519  .03202475               R-squared     =  0.9077
---------+------------------------------               Adj R-squared =  0.9042
   Total |  180.154268   539  .334237975               Root MSE      =  .17895

------------------------------------------------------------------------------
 lcrmrte |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 lprbarr |  -.1746053   .0238458     -7.322   0.000      -.2214516   -.1277591
lprbconv |  -.1337714   .0169096     -7.911   0.000       -.166991   -.1005518
lprbpris |  -.0195318   .0352873     -0.554   0.580      -.0888553    .0497918
 lavgsen |  -.1108926   .0311719     -3.557   0.000      -.1721313    -.049654
  lpolpc |   .1050704   .0172627      6.087   0.000        .071157    .1389838
     d83 |  -.0729231   .0286922     -2.542   0.011      -.1292903   -.0165559
     d84 |  -.0652494   .0287165     -2.272   0.023      -.1216644   -.0088345
```

```
     d85 |  -.0258059   .0326156     -0.791   0.429      -.0898807     .038269
     d86 |   .0263763   .0371746      0.710   0.478      -.0466549    .0994076
     d87 |   .0465632   .0418004      1.114   0.266      -.0355555    .1286819
lcrmrt_1 |   .8087768   .0208067     38.871   0.000        .767901    .8496525
   lwcon |  -.0283133   .0392516     -0.721   0.471      -.1054249    .0487983
   lwtuc |  -.0034567   .0223995     -0.154   0.877      -.0474615    .0405482
   lwtrd |   .0121236   .0439875      0.276   0.783      -.0742918     .098539
   lwfir |   .0296003   .0318995      0.928   0.354      -.0330676    .0922683
   lwser |    .012903   .0221872      0.582   0.561      -.0306847    .0564908
   lwmfg |  -.0409046   .0389325     -1.051   0.294      -.1173893    .0355801
   lwfed |   .1070534   .0798526      1.341   0.181      -.0498207    .2639275
   lwsta |  -.0903894   .0660699     -1.368   0.172      -.2201867     .039408
   lwloc |   .0961124   .1003172      0.958   0.338      -.1009652      .29319
   _cons |  -.6438061   .6335887     -1.016   0.310       -1.88852    .6009076
------------------------------------------------------------------------------

. test lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc

 ( 1)  lwcon = 0.0
 ( 2)  lwtuc = 0.0
 ( 3)  lwtrd = 0.0
 ( 4)  lwfir = 0.0
 ( 5)  lwser = 0.0
 ( 6)  lwmfg = 0.0
 ( 7)  lwfed = 0.0
 ( 8)  lwsta = 0.0
 ( 9)  lwloc = 0.0

       F(  9,   519) =    0.85
            Prob > F =    0.5663
```

CHAPTER 8

8.1. Letting $Q(\mathbf{b})$ denote the objective function in (8.23), it follows from multivariable calculus that

$$\frac{\partial Q(\mathbf{b})'}{\partial \mathbf{b}} = -2\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)'\hat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\mathbf{b})\right).$$

Evaluating the derivative at the solution $\hat{\boldsymbol{\beta}}$ gives

$$\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)'\hat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})\right) = \mathbf{0}.$$

In terms of full data matrices, we can write, after simple algebra,

$$(\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}\mathbf{Z}'\mathbf{X})\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}\mathbf{Z}'\mathbf{Y}).$$

Solving for $\hat{\boldsymbol{\beta}}$ gives (8.24).

8.3. First, we can always write $\mathbf{x}$ as its linear projection plus an error: $\mathbf{x} = \mathbf{x}^* + \mathbf{e}$, where $\mathbf{x}^* = \mathbf{z}\boldsymbol{\Pi}$ and $\mathrm{E}(\mathbf{z}'\mathbf{e}) = \mathbf{0}$. Therefore, $\mathrm{E}(\mathbf{z}'\mathbf{x}) = \mathrm{E}(\mathbf{z}'\mathbf{x}^*)$, which verifies the first part of the hint. To verify the second step, let $\mathbf{h} \equiv \mathbf{h}(\mathbf{z})$, and write the linear projection as

$$\mathrm{L}(\mathbf{y}|\mathbf{z},\mathbf{h}) = \mathbf{z}\boldsymbol{\Pi}_1 + \mathbf{h}\boldsymbol{\Pi}_2,$$

where $\boldsymbol{\Pi}_1$ is $M \times K$ and $\boldsymbol{\Pi}_2$ is $Q \times K$. Then we must show that $\boldsymbol{\Pi}_2 = \mathbf{0}$. But, from the two-step projection theorem (see Property LP.7 in Chapter 2),

$$\boldsymbol{\Pi}_2 = [\mathrm{E}(\mathbf{s}'\mathbf{s})]^{-1}\mathrm{E}(\mathbf{s}'\mathbf{r}), \text{ where } \mathbf{s} \equiv \mathbf{h} - \mathrm{L}(\mathbf{h}|\mathbf{z}) \text{ and } \mathbf{r} \equiv \mathbf{x} - \mathrm{L}(\mathbf{x}|\mathbf{z}).$$

Now, by the assumption that $\mathrm{E}(\mathbf{x}|\mathbf{z}) = \mathrm{L}(\mathbf{x}|\mathbf{z})$, $\mathbf{r}$ is also equal to $\mathbf{x} - \mathrm{E}(\mathbf{x}|\mathbf{z})$. Therefore, $\mathrm{E}(\mathbf{r}|\mathbf{z}) = \mathbf{0}$, and so $\mathbf{r}$ is uncorrelated with all functions of $\mathbf{z}$. But $\mathbf{s}$ is simply a function of $\mathbf{z}$ since $\mathbf{h} \equiv \mathbf{h}(\mathbf{z})$. Therefore, $\mathrm{E}(\mathbf{s}'\mathbf{r}) = \mathbf{0}$, and this shows that $\boldsymbol{\Pi}_2 = \mathbf{0}$.

8.5. This follows directly from the hint. Straightforward matrix algebra shows that $(\mathbf{C}'\boldsymbol{\Lambda}^{-1}\mathbf{C}) - (\mathbf{C}'\mathbf{W}\mathbf{C})(\mathbf{C}'\mathbf{W}\boldsymbol{\Lambda}\mathbf{W}\mathbf{C})^{-1}(\mathbf{C}'\mathbf{W}\mathbf{C})$ can be written as

$$\mathbf{C}'\boldsymbol{\Lambda}^{-1/2}[\mathbf{I}_L - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\boldsymbol{\Lambda}^{-1/2}\mathbf{C},$$

where $\mathbf{D} \equiv \boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C}$. Since this is a matrix quadratic form in the $L \times L$ symmetric, idempotent matrix $\mathbf{I}_L - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$, it is necessarily itself positive semi-definite.

8.7. When $\hat{\boldsymbol{\Omega}}$ is diagonal and $\mathbf{Z}_i$ has the form in (8.15), $\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\boldsymbol{\Omega}}\mathbf{Z}_i = \mathbf{Z}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}})\mathbf{Z}$ is a block diagonal matrix with g^{th} block $\hat{\sigma}_g^2\left(\sum_{i=1}^{N}\mathbf{z}_{ig}'\mathbf{z}_{ig}\right) \equiv \hat{\sigma}_g^2\mathbf{Z}_g'\mathbf{Z}_g$, where $\mathbf{Z}_g$ denotes the $N \times L_g$ matrix of instruments for the g^{th} equation. Further, $\mathbf{Z}'\mathbf{X}$ is block diagonal with g^{th} block $\mathbf{Z}_g'\mathbf{X}_g$. Using these facts, it is now straightforward to show that the 3SLS estimator consists of $[\mathbf{X}_g'\mathbf{Z}_g(\mathbf{Z}_g'\mathbf{Z}_g)^{-1}\mathbf{Z}_g'\mathbf{X}_g]^{-1}\mathbf{X}_g'\mathbf{Z}_g(\mathbf{Z}_g'\mathbf{Z}_g)^{-1}\mathbf{Z}_g'\mathbf{Y}_g$ stacked from $g = 1,\ldots,G$. This is just the system 2SLS estimator or, equivalently, 2SLS equation-by-equation.

8.9. The optimal instruments are given in Theorem 8.5, with $G = 1$:

$$\mathbf{z}_i^* = [\omega(\mathbf{z}_i)]^{-1}\mathrm{E}(\mathbf{x}_i|\mathbf{z}_i),\ \omega(\mathbf{z}_i) = \mathrm{E}(u_i^2|\mathbf{z}_i).$$

If $\mathrm{E}(u_i^2|\mathbf{z}_i) = \sigma^2$ and $\mathrm{E}(\mathbf{x}_i|\mathbf{z}_i) = \mathbf{z}_i\boldsymbol{\Pi}$, the the optimal instruments are $\sigma^{-2}\mathbf{z}_i\boldsymbol{\Pi}$. The constant multiple σ^{-2} clearly has no effect on the optimal IV estimator, so the optimal instruments are $\mathbf{z}_i\boldsymbol{\Pi}$. These are the optimal IVs underlying 2SLS, except that $\boldsymbol{\Pi}$ is replaced with its $\sqrt{N}$-consistent OLS estimator. The 2SLS estimator has the same asymptotic variance whether $\boldsymbol{\Pi}$ or $\hat{\boldsymbol{\Pi}}$ is used, and so 2SLS is asymptotically efficient.

If $\mathrm{E}(u|\mathbf{x}) = 0$ and $\mathrm{E}(u^2|\mathbf{x}) = \sigma^2$ then the optimal instruments are $\sigma^{-2}\mathrm{E}(\mathbf{x}|\mathbf{x}) = \sigma^{-2}\mathbf{x}$, and this leads to the OLS estimator.

8.11. a. This is a simple application of Theorem 8.5 when $G = 1$. Without the i subscript, $\mathbf{x}_1 = (\mathbf{z}_1,y_2)$ and so $\mathrm{E}(\mathbf{x}_1|\mathbf{z}) = [\mathbf{z}_1,\mathrm{E}(y_2|\mathbf{z})]$. Further, $\boldsymbol{\Omega}(\mathbf{z}) = \mathrm{Var}(u_1|\mathbf{z}) = \sigma_1^2$. It follows that the optimal instruments are $(1/\sigma_1^2)[\mathbf{z}_1,\mathrm{E}(y_2|\mathbf{z})]$. Dropping the division by σ_1^2 clearly does not affect the optimal instruments.

b. If y_2 is binary then $\mathrm{E}(y_2|\mathbf{z}) = \mathrm{P}(y_2 = 1|\mathbf{z}) = F(\mathbf{z})$, and so the optimal IVs are $[\mathbf{z}_1,F(\mathbf{z})]$.

CHAPTER 9

9.1. a. No. What causal inference could one draw from this? We may be interested in the tradeoff between wages and benefits, but then either of these can be taken as the dependent variable and the analysis would be by OLS. Of course, if we have omitted some important factors or have a measurement error problem, OLS could be inconsistent for estimating the tradeoff. But it is not a simultaneity problem.

b. Yes. We can certainly think of an exogenous change in law enforcement expenditures causing a reduction in crime, and we are certainly interested in such thought experiments. If we could do the appropriate experiment, where expenditures are assigned randomly across cities, then we could estimate the crime equation by OLS. (In fact, we could use a simple regression analysis.) The simultaneous equations model recognizes that cities choose law enforcement expenditures in part on what they expect the crime rate to be. An SEM is a convenient way to allow expenditures to depend on unobservables (to the econometrician) that affect crime.

c. No. These are both choice variables of the firm, and the parameters in a two-equation system modeling one in terms of the other, and vice versa, have no economic meaning. If we want to know how a change in the price of foreign technology affects foreign technology (FT) purchases, why would we want to hold fixed R&D spending? Clearly FT purchases and R&D spending are simultaneously chosen, but we should use a SUR model where neither is an explanatory variable in the other’s equation.

d. Yes. We we can certainly be interested in the causal effect of alcohol consumption on productivity, and therefore wage. One’s hourly wage is determined by the demand for skills; alcohol consumption is determined by individual behavior.

e. No. These are choice variables by the same household. It makes no sense to think about how exogenous changes in one would affect the other. Further, suppose that we look at the effects of changes in local property tax rates. We would not want to hold fixed family saving and then measure the effect of changing property taxes on housing expenditures. When the property tax changes, a family will generally adjust expenditure in all categories. A SUR system with property tax as an explanatory variable seems to be the appropriate model.

f. No. These are both chosen by the firm, presumably to maximize profits. It makes no sense to hold advertising expenditures fixed while looking at how other variables affect price markup.

9.3. a. We can apply part b of Problem 9.2. First, the only variable excluded from the *support* equation is the variable *mremarr*; since the *support* equation contains one endogenous variable, this equation is identified if and only if $\delta_{21} \neq 0$. This ensures that there is an exogenous variable shifting the mother's reaction function that does not also shift the father's reaction function.

The *visits* equation is identified if and only if at least one of *finc* and *fremarr* actually appears in the *support* equation; that is, we need $\delta_{11} \neq 0$ or $\delta_{13} \neq 0$.

b. Each equation can be estimated by 2SLS using instruments 1, *finc*, *fremarr*, *dist*, *mremarr*.

c. First, obtain the reduced form for *visits*:

$$visits = \pi_{20} + \pi_{21}finc + \pi_{22}fremarr + \pi_{23}dist + \pi_{24}mremarr + v_2.$$

Estimate this equation by OLS, and save the residuals, $\hat{v}_2$. Then, run the OLS regression

$$support \text{ on } 1, visits, finc, fremarr, dist, \hat{v}_2$$

and do a (heteroskedasticity-robust) *t* test that the coefficient on $\hat{v}_2$ is zero. If this test rejects we conclude that *visits* is in fact endogenous in the *support* equation.

d. There is one overidentifying restriction in the *visits* equation, assuming that δ_{11} and δ_{12} are both different from zero. Assuming homoskedasticity of u_2, the easiest way to test the overidentifying restriction is to first estimate the *visits* equation by 2SLS, as in part b. Let $\hat{u}_2$ be the 2SLS residuals. Then, run the auxiliary regression

$$\hat{u}_2 \text{ on } 1, finc, fremarr, dist, mremarr;$$

the sample size times the usual *R*-squared from this regression is distributed asymptotically as χ_1^2 under the null hypothesis that all instruments are exogenous.

A heteroskedasticity-robust test is also easy to obtain. Let $\widehat{support}$ denote the fitted values from the reduced form regression for *support*. Next, regress *finc* (or *fremarr*) on $\widehat{support}$, *mremarr*, *dist*, and save the residuals, say $\hat{r}_1$. Then, run the simple regression (without intercept) of 1 on $\hat{u}_2\hat{r}_1$; $N - \text{SSR}_0$ from this regression is asymptotically χ_1^2 under H_0. (SSR_0 is just the usual sum of squared residuals.)

9.5. a. Let $\boldsymbol{\beta}_1$ denote the 7×1 vector of parameters in the first equation with only the normalization restriction imposed:

$$\boldsymbol{\beta}_1' = (-1, \gamma_{12}, \gamma_{13}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}).$$

The restrictions $\delta_{12} = 0$ and $\delta_{13} + \delta_{14} = 1$ are obtained by choosing

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Because $\mathbf{R}_1$ has two rows, and $G - 1 = 2$, the order condition is satisfied. Now, we need to check the rank condition. Letting $\mathbf{B}$ denote the 7×3 matrix of all structural parameters with only the three normalizations, straightforward matrix multiplication gives

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{13} + \delta_{14} - 1 & \delta_{23} + \delta_{24} - \gamma_{21} & \delta_{33} + \delta_{34} - \gamma_{31} \end{pmatrix}.$$

By definition of the constraints on the first equation, the first column of $\mathbf{R}_1\mathbf{B}$ is zero. Next, we use the constraints in the remainder of the system to get the expression for $\mathbf{R}_1\mathbf{B}$ with all information imposed. But $\gamma_{23} = 0$, $\delta_{22} = 0$, $\delta_{23} = 0$, $\delta_{24} = 0$, $\gamma_{31} = 0$, and $\gamma_{32} = 0$, and so $\mathbf{R}_1\mathbf{B}$ becomes

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} 0 & 0 & \delta_{32} \\ 0 & -\gamma_{21} & \delta_{33} + \delta_{34} - \gamma_{31} \end{pmatrix}.$$

Identification requires $\gamma_{21} \neq 0$ and $\delta_{32} \neq 0$.

b. It is easy to see how to estimate the first equation under the given assumptions. Set $\delta_{14} = 1 - \delta_{13}$ and plug this into the equation. After simple algebra we get

$$y_1 - z_4 = \gamma_{12}y_2 + \gamma_{13}y_3 + \delta_{11}z_1 + \delta_{13}(z_3 - z_4) + u_1.$$

This equation can be estimated by 2SLS using instruments (z_1,z_2,z_3,z_4). Note that, if we just count instruments, there are just enough instruments to estimate this equation.

9.7. a. Because *alcohol* and *educ* are endogenous in the first equation, we need at least two elements in $\mathbf{z}_{(2)}$ and/or $\mathbf{z}_{(3)}$ that are not also in $\mathbf{z}_{(1)}$. Ideally, we have at least one such element in $\mathbf{z}_{(2)}$ and at least one such element in $\mathbf{z}_{(3)}$.

b. Let $\mathbf{z}$ denote all nonredundant exogenous variables in the system. Then use these as instruments in a 2SLS analysis.

c. The matrix of instruments for each *i* is

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{z}_i, educ_i) & \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_i \end{pmatrix}.$$

d. $\mathbf{z}_{(3)} = \mathbf{z}$. That is, we should not make any exclusion restrictions in the reduced form for *educ*.

9.9. a. Here is my Stata output for the 3SLS estimation of (9.28) and (9.29):

```
. reg3 (hours lwage educ age kidslt6 kidsge6 nwifeinc) (lwage hours educ exper
expersq)

Three-stage least squares regression
----------------------------------------------------------------------
Equation          Obs  Parms        RMSE    "R-sq"       chi2        P
----------------------------------------------------------------------
hours             428      6    1368.362   -2.1145   34.53608   0.0000
lwage             428      4    .6892584    0.0895   79.87188   0.0000
----------------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
       lwage |   1676.933    431.169     3.89   0.000     831.8577    2522.009
        educ |  -205.0267   51.84729    -3.95   0.000    -306.6455   -103.4078
         age |  -12.28121   8.261529    -1.49   0.137    -28.47351    3.911094
     kidslt6 |  -200.5673   134.2685    -1.49   0.135    -463.7287    62.59414
     kidsge6 |  -48.63986   35.95137    -1.35   0.176    -119.1032    21.82352
    nwifeinc |   .3678943   3.451518     0.11   0.915    -6.396957    7.132745
       _cons |   2504.799   535.8919     4.67   0.000      1454.47    3555.128
-------------+----------------------------------------------------------------
lwage        |
       hours |    .000201   .0002109     0.95   0.340    -.0002123    .0006143
        educ |   .1129699   .0151452     7.46   0.000     .0832858    .1426539
       exper |   .0208906   .0142782     1.46   0.143    -.0070942    .0488753
```

```
     expersq |  -.0002943   .0002614    -1.13   0.260    -.0008066     .000218
       _cons |  -.7051103   .3045904    -2.31   0.021    -1.302097   -.1081241
------------------------------------------------------------------------------
Endogenous variables:  hours lwage
Exogenous variables:   educ age kidslt6 kidsge6 nwifeinc exper expersq
------------------------------------------------------------------------------
```

b. To be added. Unfortunately, I know of no econometrics packages that conveniently allow system estimation using different instruments for different equations.

9.11. a. Since z_2 and z_3 are both omitted from the first equation, we just need $\delta_{22} \neq 0$ or $\delta_{23} \neq 0$ (or both, of course). The second equation is identified if and only if $\delta_{11} \neq 0$.

b. After substitution and straightforward algebra, it can be seen that $\pi_{11} = \delta_{11}/(1 - \gamma_{12}\gamma_{21})$.

c. We can estimate the system by 3SLS; for the second equation, this is identical to 2SLS since it is just identified. Or, we could just use 2SLS on each equation. Given $\hat{\delta}_{11}$, $\hat{\gamma}_{12}$, and $\hat{\gamma}_{21}$, we would form $\hat{\pi}_{11} = \hat{\delta}_{11}/(1 - \hat{\gamma}_{12}\hat{\gamma}_{21})$.

d. Whether we estimate the parameters by 2SLS or 3SLS, we will generally inconsistently estimate δ_{11} and γ_{12}. (Since we are estimating the second equation by 2SLS, we will still consistently estimate γ_{21} provided we have not misspecified this equation.) So our estimate of $\pi_{11} = \partial\mathrm{E}(y_2|\mathbf{z})/\partial z_1$ will be inconsistent in any case.

e. We can just estimate the reduced form $\mathrm{E}(y_2|z_1,z_2,z_3)$ by ordinary least squares.

f. Consistency of OLS for π_{11} does not hinge on the validity of the exclusion restrictions in the structural model, whereas using an SEM does. Of course, if the SEM is correctly specified, we obtain a more efficient estimator of the reduced form parameters by imposing the restrictions in estimating π_{11}.

9.13. a. The first equation is identified if, and only if, $\delta_{22} \neq 0$. (This is the rank condition.)

b. Here is my Stata output:

```
. reg open lpcinc lland

  Source |       SS       df       MS                  Number of obs =     114
---------+------------------------------               F(  2,   111) =   45.17
   Model |  28606.1936     2  14303.0968               Prob > F      =  0.0000
Residual |  35151.7966   111  316.682852               R-squared     =  0.4487
---------+------------------------------               Adj R-squared =  0.4387
   Total |  63757.9902   113  564.230002               Root MSE      =  17.796

------------------------------------------------------------------------------
    open |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lpcinc |   .5464812    1.49324      0.366   0.715      -2.412473    3.505435
   lland |  -7.567103   .8142162     -9.294   0.000      -9.180527   -5.953679
   _cons |   117.0845    15.8483      7.388   0.000       85.68006     148.489
------------------------------------------------------------------------------
```

This shows that log(*land*) is very statistically significant in the RF for *open*. Smaller countries are more open.

c. Here is my Stata output. First, 2SLS, then OLS:

```
. reg inf open lpcinc (lland lpcinc)
                                                                        (2SLS)
  Source |       SS       df       MS                  Number of obs =     114
---------+------------------------------               F(  2,   111) =    2.79
   Model |  2009.22775     2  1004.61387               Prob > F      =  0.0657
Residual |   63064.194   111  568.145892               R-squared     =  0.0309
---------+------------------------------               Adj R-squared =  0.0134
   Total |  65073.4217   113  575.870989               Root MSE      =  23.836

------------------------------------------------------------------------------
     inf |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    open |  -.3374871   .1441212     -2.342   0.021      -.6230728   -.0519014
  lpcinc |   .3758247   2.015081      0.187   0.852      -3.617192    4.368841
   _cons |   26.89934    15.4012      1.747   0.083       -3.61916    57.41783
```

```
. reg inf open lpcinc

  Source |       SS       df       MS                  Number of obs =     114
---------+------------------------------               F(  2,   111) =    2.63
   Model |  2945.92812     2  1472.96406               Prob > F      =  0.0764
Residual |  62127.4936   111   559.70715               R-squared     =  0.0453
---------+------------------------------               Adj R-squared =  0.0281
   Total |  65073.4217   113  575.870989               Root MSE      =  23.658

------------------------------------------------------------------------------
     inf |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    open |  -.2150695   .0946289     -2.273   0.025       -.402583    -.027556
  lpcinc |   .0175683   1.975267      0.009   0.993      -3.896555    3.931692
   _cons |   25.10403   15.20522      1.651   0.102      -5.026122    55.23419
------------------------------------------------------------------------------
```

The 2SLS estimate is notably larger in magnitude. Not surprisingly, it also has a larger standard error. You might want to test to see if *open* is endogenous.

d. If we add $\gamma_{13}open^2$ to the equation, we need an IV for it. Since log(*land*) is partially correlated with *open*, $[\log(land)]^2$ is a natural candidate. A regression of $open^2$ on log(*land*), $[\log(land)]^2$, and log(*pcinc*) gives a heteroskedasticity-robust *t* statistic on $[\log(land)]^2$ of about 2. This is borderline, but we will go ahead. The Stata output for 2SLS is

```
. gen opensq = open^2

. gen llandsq = lland^2

. reg inf open opensq lpcinc (lland llandsq lpcinc)
                                                                       (2SLS)
  Source |       SS       df       MS                  Number of obs =     114
---------+------------------------------               F(  3,   110) =    2.09
   Model | -414.331026     3 -138.110342               Prob > F      =  0.1060
Residual |  65487.7527   110  595.343207               R-squared     =       .
---------+------------------------------               Adj R-squared =       .
   Total |  65073.4217   113  575.870989               Root MSE      =   24.40

------------------------------------------------------------------------------
     inf |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
```

```
    open |  -1.198637   .6205699     -1.932   0.056      -2.428461    .0311868
  opensq |   .0075781   .0049828      1.521   0.131      -.0022966    .0174527
  lpcinc |   .5066092   2.069134      0.245   0.807      -3.593929    4.607147
   _cons |   43.17124   19.36141      2.230   0.028       4.801467    81.54102
------------------------------------------------------------------------------
```

The squared term indicates that the impact of *open* on *inf* diminishes; the estimate would be significant at about the 6.5% level against a one-sided alternative.

e. Here is the Stata output for implementing the method described in the problem:

```
. reg open lpcinc lland

      Source |       SS       df       MS              Number of obs =     114
-------------+------------------------------           F(  2,   111) =   45.17
       Model |  28606.1936     2  14303.0968           Prob > F      =  0.0000
    Residual |  35151.7966   111  316.682852           R-squared     =  0.4487
-------------+------------------------------           Adj R-squared =  0.4387
       Total |  63757.9902   113  564.230002           Root MSE      =  17.796

------------------------------------------------------------------------------
        open |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lpcinc |   .5464812    1.49324     0.37   0.715    -2.412473    3.505435
       lland |  -7.567103   .8142162    -9.29   0.000    -9.180527   -5.953679
       _cons |   117.0845    15.8483     7.39   0.000     85.68006     148.489
------------------------------------------------------------------------------

. predict openh
(option xb assumed; fitted values)

. gen openhsq = openh^2

. reg inf openh openhsq lpcinc

      Source |       SS       df       MS              Number of obs =     114
-------------+------------------------------           F(  3,   110) =    2.24
       Model |  3743.18411     3  1247.72804           Prob > F      =  0.0879
    Residual |  61330.2376   110  557.547615           R-squared     =  0.0575
-------------+------------------------------           Adj R-squared =  0.0318
       Total |  65073.4217   113  575.870989           Root MSE      =  23.612

------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
        openh |  -.8648092   .5394132    -1.60   0.112    -1.933799     .204181
      openhsq |   .0060502   .0059682     1.01   0.313    -.0057774    .0178777
       lpcinc |   .0412172   2.023302     0.02   0.984    -3.968493    4.050927
        _cons |   39.17831   19.48041     2.01   0.047     .5727026    77.78391
------------------------------------------------------------------------------
```

Qualitatively, the results are similar to the correct IV method from part d. If $\gamma_{13} = 0$, E(*open*|*lpcinc*,*lland*) is linear and, as shown in Problem 9.12, both methods are consistent. But the forbidden regression implemented in this part is uncessary, less robust, and we cannot trust the standard errors, anyway.

CHAPTER 10

10.1. a. Since investment is likely to be affected by macroeconomic factors, it is important to allow for these by including separate time intercepts; this is done by using $T - 1$ time period dummies.

b. Putting the unobserved effect c_i in the equation is a simple way to account for time-constant features of a county that affect investment and might also be correlated with the tax variable. Something like "average" county economic climate, which affects investment, could easily be correlated with tax rates because tax rates are, at least to a certain extent, selected by state and local officials. If only a cross section were available, we would have to find an instrument for the tax variable that is uncorrelated with c_i and correlated with the tax rate. This is often a difficult task.

c. Standard investment theories suggest that, ceteris paribus, larger marginal tax rates decrease investment.

d. I would start with a fixed effects analysis to allow arbitrary correlation between all time-varying explanatory variables and c_i. (Actually, doing pooled OLS is a useful initial exercise; these results can be compared with those from an FE analysis). Such an analysis assumes strict exogeneity of $\mathbf{z}_{it}$, tax_{it}, and $disaster_{it}$ in the sense that these are uncorrelated with the errors u_{is} for all t and s.

I have no strong intuition for the likely serial correlation properties of the $\{u_{it}\}$. These might have little serial correlation because we have allowed for c_i, in which case I would use standard fixed effects. However, it seems more likely that the u_{it} are positively autocorrelated, in which case I might use first differencing instead. In either case, I would compute the fully robust standard errors along with the usual ones. Remember, with first-differencing it is easy to test whether the changes Δu_{it} are serially uncorrelated.

e. If tax_{it} and $disaster_{it}$ do not have lagged effects on investment, then the only possible violation of the strict exogeneity assumption is if future values of these variables are correlated with u_{it}. It is safe to say that this is not a worry for the disaster variable: presumably, future natural disasters are not determined by past investment. On the other hand, state officials might look at the levels of past investment in determining future tax policy, especially if there is a target level of tax revenue the officials are are trying to achieve. This could be similar to setting property tax rates: sometimes property tax rates are set depending on recent housing values, since a larger base means a smaller rate can achieve the same amount of revenue. Given that we allow tax_{it} to be correlated with c_i, this might not be much of a problem. But it cannot be ruled out ahead of time.

10.3. a. Let $\bar{\mathbf{x}}_i = (\mathbf{x}_{i1} + \mathbf{x}_{i2})/2$, $\bar{y}_i = (y_{i1} + y_{i2})/2$, $\ddot{\mathbf{x}}_{i1} = \mathbf{x}_{i1} - \bar{\mathbf{x}}_i$,

$\ddot{\mathbf{x}}_{i2} = \mathbf{x}_{i2} - \bar{\mathbf{x}}_i$, and similarly for $\ddot{y}_{i1}$ and $\ddot{y}_{i2}$. For $T = 2$ the fixed effects estimator can be written as

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N}(\ddot{\mathbf{x}}_{i1}'\ddot{\mathbf{x}}_{i1} + \ddot{\mathbf{x}}_{i2}'\ddot{\mathbf{x}}_{i2})\right)^{-1}\left(\sum_{i=1}^{N}(\ddot{\mathbf{x}}_{i1}'\ddot{y}_{i1} + \ddot{\mathbf{x}}_{i2}'\ddot{y}_{i2})\right).$$

Now, by simple algebra,

$$\ddot{\mathbf{x}}_{i1} = (\mathbf{x}_{i1} - \mathbf{x}_{i2})/2 = -\Delta\mathbf{x}_i/2$$
$$\ddot{\mathbf{x}}_{i2} = (\mathbf{x}_{i2} - \mathbf{x}_{i1})/2 = \Delta\mathbf{x}_i/2$$
$$\ddot{y}_{i1} = (y_{i1} - y_{i2})/2 = -\Delta y_i/2$$
$$\ddot{y}_{i2} = (y_{i2} - y_{i1})/2 = \Delta y_i/2.$$

Therefore,

$$\ddot{\mathbf{x}}_{i1}'\ddot{\mathbf{x}}_{i1} + \ddot{\mathbf{x}}_{i2}'\ddot{\mathbf{x}}_{i2} = \Delta\mathbf{x}_i'\Delta\mathbf{x}_i/4 + \Delta\mathbf{x}_i'\Delta\mathbf{x}_i/4 = \Delta\mathbf{x}_i'\Delta\mathbf{x}_i/2$$
$$\ddot{\mathbf{x}}_{i1}'\ddot{y}_{i1} + \ddot{\mathbf{x}}_{i2}'\ddot{y}_{i2} = \Delta\mathbf{x}_i'\Delta y_i/4 + \Delta\mathbf{x}_i'\Delta y_i/4 = \Delta\mathbf{x}_i'\Delta y_i/2,$$

and so

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N}\Delta\mathbf{x}_i'\Delta\mathbf{x}_i/2\right)^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{x}_i'\Delta y_i/2\right)$$
$$= \left(\sum_{i=1}^{N}\Delta\mathbf{x}_i'\Delta\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{x}_i'\Delta y_i\right) = \hat{\boldsymbol{\beta}}_{FD}.$$

b. Let $\hat{u}_{i1} = \ddot{y}_{i1} - \ddot{\mathbf{x}}_{i1}\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{u}_{i2} = \ddot{y}_{i2} - \ddot{\mathbf{x}}_{i2}\hat{\boldsymbol{\beta}}_{FE}$ be the fixed effects residuals for the two time periods for cross section observation i. Since $\hat{\boldsymbol{\beta}}_{FE} = \hat{\boldsymbol{\beta}}_{FD}$, and using the representations in (4.1′), we have

$$\hat{u}_{i1} = -\Delta y_i/2 - (-\Delta\mathbf{x}_i/2)\hat{\boldsymbol{\beta}}_{FD} = -(\Delta y_i - \Delta\mathbf{x}_i\hat{\boldsymbol{\beta}}_{FD})/2 \equiv -\hat{e}_i/2$$
$$\hat{u}_{i2} = \Delta y_i/2 - (\Delta\mathbf{x}_i/2)\hat{\boldsymbol{\beta}}_{FD} = (\Delta y_i - \Delta\mathbf{x}_i\hat{\boldsymbol{\beta}}_{FD})/2 \equiv \hat{e}_i/2,$$

where $\hat{e}_i \equiv \Delta y_i - \Delta\mathbf{x}_i\hat{\boldsymbol{\beta}}_{FD}$ are the first difference residuals, $i = 1,2,\ldots,N$. Therefore,

$$\sum_{i=1}^{N}(\hat{u}_{i1}^2 + \hat{u}_{i2}^2) = (1/2)\sum_{i=1}^{N}\hat{e}_i^2.$$

This shows that the sum of squared residuals from the fixed effects regression is exactly one have the sum of squared residuals from the first difference regression. Since we know the variance estimate for fixed effects is the SSR divided by $N - K$ (when $T = 2$), and the variance estimate for first difference is the SSR divided by $N - K$, the error variance from fixed effects is always half the size as the error variance for first difference estimation, that is, $\hat{\sigma}_u^2 = \hat{\sigma}_e^2/2$ (contrary to what the problem asks you so show). What I wanted you to show is that the variance matrix estimates of $\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{\boldsymbol{\beta}}_{FD}$ are identical. This is easy since the variance matrix estimate for fixed effects is

$$\hat{\sigma}_u^2\left(\sum_{i=1}^{N}(\ddot{\mathbf{x}}_{i1}'\ddot{\mathbf{x}}_{i1} + \ddot{\mathbf{x}}_{i2}'\ddot{\mathbf{x}}_{i2})\right)^{-1} = (\hat{\sigma}_e^2/2)\left(\sum_{i=1}^{N}\Delta\mathbf{x}_i'\Delta\mathbf{x}_i/2\right)^{-1} = \hat{\sigma}_e^2\left(\sum_{i=1}^{N}\Delta\mathbf{x}_i'\Delta\mathbf{x}_i\right)^{-1},$$

which is the variance matrix estimator for first difference. Thus, the standard errors, and in fact all other test statistics (F statistics) will be numerically identical using the two approaches.

10.5. a. Write $\mathbf{v}_i\mathbf{v}_i' = c_i^2\mathbf{j}_T\mathbf{j}_T' + \mathbf{u}_i\mathbf{u}_i' + \mathbf{j}_T(c_i\mathbf{u}_i') + (c_i\mathbf{u}_i)\mathbf{j}_T'$. Under RE.1, $\mathrm{E}(\mathbf{u}_i|\mathbf{x}_i,c_i) = \mathbf{0}$, which implies that $\mathrm{E}[(c_i\mathbf{u}_i')|\mathbf{x}_i) = \mathbf{0}$ by interated expecations. Under RE.3a, $\mathrm{E}(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i,c_i) = \sigma_u^2\mathbf{I}_T$, which implies that $\mathrm{E}(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i) = \sigma_u^2\mathbf{I}_T$ (again, by iterated expectations). Therefore,

$$\mathrm{E}(\mathbf{v}_i\mathbf{v}_i'|\mathbf{x}_i) = \mathrm{E}(c_i^2|\mathbf{x}_i)\mathbf{j}_T\mathbf{j}_T' + \mathrm{E}(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i) = h(\mathbf{x}_i)\mathbf{j}_T\mathbf{j}_T' + \sigma_u^2\mathbf{I}_T,$$

where $h(\mathbf{x}_i) \equiv \mathrm{Var}(c_i|\mathbf{x}_i) = \mathrm{E}(c_i^2|\mathbf{x}_i)$ (by RE.1b). This shows that the conditional variance matrix of $\mathbf{v}_i$ given $\mathbf{x}_i$ has the same covariance for all $t \neq s$, $h(\mathbf{x}_i)$, and the same variance for all t, $h(\mathbf{x}_i) + \sigma_u^2$. Therefore, while the variances and covariances depend on $\mathbf{x}_i$ in general, they do not depend on time separately.

b. The RE estimator is still consistent and $\sqrt{N}$-asymptotically normal without assumption RE.3b, but the usual random effects variance estimator of $\hat{\boldsymbol{\beta}}_{RE}$ is no longer valid because $\mathrm{E}(\mathbf{v}_i\mathbf{v}_i'|\mathbf{x}_i)$ does not have the form (10.30) (because it depends on $\mathbf{x}_i$). The robust variance matrix estimator given in (7.49) should be used in obtaining standard errors or Wald statistics.

10.7. I provide annotated Stata output, and I compute the nonrobust regression-based statistic from equation (11.79):

```
. * random effects estimation

. iis id

. tis term

. xtreg trmgpa spring crsgpa frstsem season sat verbmath hsperc hssize black
female, re

                                                    Random-effects GLS regression
sd(u_id)                     =  .3718544                  Number of obs =      732
sd(e_id_t)                   =  .4088283                              n =      366
sd(e_id_t + u_id)            =  .5526448                              T =        2

corr(u_id, X)                =  0 (assumed)               R-sq within   =   0.2067
                                                               between   =   0.5390
                                                               overall   =   0.4785

                                                          chi2( 10)     =   512.77
(theta = 0.3862)                                            Prob > chi2 =   0.0000

------------------------------------------------------------------------------
  trmgpa |      Coef.   Std. Err.        z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  spring |  -.0606536   .0371605     -1.632   0.103       -.1334868    .0121797
  crsgpa |   1.082365   .0930877     11.627   0.000        .8999166    1.264814
 frstsem |   .0029948   .0599542      0.050   0.960       -.1145132    .1205028
  season |  -.0440992   .0392381     -1.124   0.261       -.1210044    .0328061
     sat |   .0017052   .0001771      9.630   0.000        .0013582    .0020523
verbmath |  -.1575199     .16351     -0.963   0.335       -.4779937    .1629538
  hsperc |  -.0084622   .0012426     -6.810   0.000       -.0108977   -.0060268
  hssize |  -.0000775   .0001248     -0.621   0.534        -.000322     .000167
   black |  -.2348189   .0681573     -3.445   0.000       -.3684048   -.1012331
  female |   .3581529   .0612948      5.843   0.000        .2380173    .4782886
   _cons |   -1.73492   .3566599     -4.864   0.000        -2.43396   -1.035879
------------------------------------------------------------------------------

. * fixed effects estimation, with time-varying variables only.

. xtreg trmgpa spring crsgpa frstsem season, fe
```

```
                                           Fixed-effects (within) regression
sd(u_id)                    =   .679133         Number of obs =      732
sd(e_id_t)                  =  .4088283                     n =      366
sd(e_id_t + u_id)           =   .792693                     T =        2

corr(u_id, Xb)              =   -0.0893         R-sq within   =   0.2069
                                                     between  =   0.0333
                                                     overall  =   0.0613

                                                F(  4,   362) =    23.61
                                                     Prob > F =   0.0000

------------------------------------------------------------------------------
  trmgpa |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  spring |  -.0657817   .0391404     -1.681   0.094      -.1427528    .0111895
  crsgpa |   1.140688   .1186538      9.614   0.000       .9073506    1.374025
 frstsem |   .0128523   .0688364      0.187   0.852      -.1225172    .1482218
  season |  -.0566454   .0414748     -1.366   0.173      -.1382072    .0249165
   _cons |  -.7708056   .3305004     -2.332   0.020      -1.420747   -.1208637
------------------------------------------------------------------------------
      id |          F(365,362) =      5.399   0.000          (366 categories)

. * Obtaining the regression-based Hausman test is a bit tedious.  First,
compute the time-averages for all of the time-varying variables:

. egen atrmgpa = mean(trmgpa), by(id)

. egen aspring = mean(spring), by(id)

. egen acrsgpa = mean(crsgpa), by(id)

. egen afrstsem = mean(frstsem), by(id)

. egen aseason = mean(season), by(id)

. * Now obtain GLS transformations for both time-constant and
. * time-varying variables.  Note that lamdahat = .386.

. di 1 - .386
.614

. gen bone = .614

. gen bsat = .614*sat

. gen bvrbmth = .614*verbmath

. gen bhsperc = .614*hsperc

. gen bhssize = .614*hssize
```

```
. gen bblack = .614*black

. gen bfemale = .614*female

. gen btrmgpa = trmgpa - .386*atrmgpa

. gen bspring = spring - .386*aspring

. gen bcrsgpa = crsgpa - .386*acrsgpa

. gen bfrstsem = frstsem - .386*afrstsem

. gen bseason = season - .386*aseason

. * Check to make sure that pooled OLS on transformed data is random
. * effects.

. reg btrmgpa bone bspring bcrsgpa bfrstsem bseason bsat bvrbmth bhsperc
bhssize bblack bfemale, nocons

  Source |       SS       df       MS                  Number of obs =     732
---------+------------------------------               F( 11,   721) =  862.67
   Model |  1584.10163    11  144.009239               Prob > F      =  0.0000
Residual |  120.359125   721    .1669336               R-squared     =  0.9294
---------+------------------------------               Adj R-squared =  0.9283
   Total |  1704.46076   732   2.3284983               Root MSE      =  .40858

------------------------------------------------------------------------------
 btrmgpa |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    bone |  -1.734843   .3566396     -4.864   0.000      -2.435019   -1.034666
 bspring |   -.060651   .0371666     -1.632   0.103      -.1336187    .0123167
 bcrsgpa |   1.082336   .0930923     11.626   0.000       .8995719    1.265101
bfrstsem |   .0029868   .0599604      0.050   0.960       -.114731    .1207046
 bseason |  -.0440905   .0392441     -1.123   0.262      -.1211368    .0329558
    bsat |   .0017052    .000177      9.632   0.000       .0013577    .0020528
 bvrbmth |  -.1575166   .1634784     -0.964   0.336      -.4784672     .163434
 bhsperc |  -.0084622   .0012424     -6.811   0.000      -.0109013   -.0060231
 bhssize |  -.0000775   .0001247     -0.621   0.535      -.0003224    .0001674
  bblack |  -.2348204   .0681441     -3.446   0.000      -.3686049   -.1010359
 bfemale |   .3581524   .0612839      5.844   0.000       .2378363    .4784686
------------------------------------------------------------------------------

. * These are the RE estimates, subject to rounding error.

. * Now add the time averages of the variables that change across i and t
. * to perform the Hausman test:

. reg btrmgpa bone bspring bcrsgpa bfrstsem bseason bsat bvrbmth bhsperc
bhssize bblack bfemale acrsgpa afrstsem aseason, nocons
```

```
  Source |       SS       df       MS                  Number of obs =     732
---------+------------------------------               F( 14,   718) =  676.85
   Model |  1584.40773    14  113.171981               Prob > F      =  0.0000
Residual |  120.053023   718  .167204767               R-squared     =  0.9296
---------+------------------------------               Adj R-squared =  0.9282
   Total |  1704.46076   732   2.3284983               Root MSE      =  .40891

------------------------------------------------------------------------------
 btrmgpa |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    bone |  -1.423761   .5182286     -2.747   0.006      -2.441186   -.4063367
 bspring |  -.0657817   .0391479     -1.680   0.093      -.1426398    .0110764
 bcrsgpa |   1.140688   .1186766      9.612   0.000       .9076934    1.373683
bfrstsem |   .0128523   .0688496      0.187   0.852      -.1223184     .148023
 bseason |  -.0566454   .0414828     -1.366   0.173      -.1380874    .0247967
    bsat |   .0016681   .0001804      9.247   0.000        .001314    .0020223
 bvrbmth |  -.1316462   .1654425     -0.796   0.426      -.4564551    .1931626
 bhsperc |  -.0084655   .0012551     -6.745   0.000      -.0109296   -.0060013
 bhssize |  -.0000783   .0001249     -0.627   0.531      -.0003236     .000167
  bblack |  -.2447934   .0685972     -3.569   0.000      -.3794684   -.1101184
 bfemale |   .3357016   .0711669      4.717   0.000       .1959815    .4754216
 acrsgpa |  -.1142992   .1234835     -0.926   0.355      -.3567312    .1281327
afrstsem |  -.0480418   .0896965     -0.536   0.592      -.2241405    .1280569
 aseason |   .0763206   .0794119      0.961   0.337      -.0795867    .2322278
------------------------------------------------------------------------------

. test acrsgpa afrstsem aseason

 ( 1)  acrsgpa = 0.0
 ( 2)  afrstsem = 0.0
 ( 3)  aseason = 0.0

       F(  3,   718) =    0.61
            Prob > F =    0.6085

. * Thus, we fail to reject the random effects assumptions even at very large
. * significance levels.
```

For comparison, the usual form of the Hausman test, which includes *spring* among the coefficients tested, gives *p*-value = .770, based on a χ^2_4 distribution (using Stata 7.0). It would have been easy to make the regression-based test robust to any violation of RE.3: add ", robust cluster(id)" to the regression command.

10.9. a. The Stata output follows. The simplest way to compute a Hausman test is to just add the time averages of all explanatory variables, excluding the dummy variables, and estimating the equation by random effects. I should have done a better job of spelling this out in the text. In other words, write

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{w}}_i\boldsymbol{\xi} + r_{it}, \quad t = 1,\ldots,T,$$

where $\mathbf{x}_{it}$ includes an overall intercept along with time dummies, as well as $\mathbf{w}_{it}$, the covariates that change across i and t. We can estimate this equation by random effects and test H_0: $\boldsymbol{\xi} = \mathbf{0}$. The actual calculation for this example is to be added.

Parts b, c, and d: To be added.

10.11. To be added.

10.13. The short answer is: Yes, we can justify this procedure with fixed T as $N \to \infty$. In particular, it produces a $\sqrt{N}$-consistent, asymptotically normal estimator of $\boldsymbol{\beta}$. Therefore, "fixed effects weighted least squares," where the weights are known functions of exogenous variables (including $\mathbf{x}_i$ and possible other covariates that do not appear in the conditional mean), is another case where "estimating" the fixed effects leads to an estimator of $\boldsymbol{\beta}$ with good properties. (As usual with fixed T, there is no sense in which we can estimate the c_i consistently.) Verifying this claim takes much more work, but it is mostly just algebra.

First, in the sum of squared residuals, we can "concentrate" the a_i out by finding $\hat{a}_i(\mathbf{b})$ as a function of $(\mathbf{x}_i,\mathbf{y}_i)$ and $\mathbf{b}$, substituting back into the sum of squared residuals, and then minimizing with respect to $\mathbf{b}$ only. Straightforward algebra gives the first order conditions for each i as

$$\sum_{t=1}^{T}(y_{it} - \hat{a}_i - \mathbf{x}_{it}\mathbf{b})/h_{it} = 0,$$

which gives

$$\hat{a}_i(\mathbf{b}) = w_i\left(\sum_{t=1}^{T} y_{it}/h_{it}\right) - w_i\left(\sum_{t=1}^{T}\mathbf{x}_{it}/h_{it}\right)\mathbf{b}$$
$$\equiv \bar{y}_i^w - \bar{\mathbf{x}}_i^w\mathbf{b},$$

where $w_i \equiv 1/\left(\sum_{t=1}^{T}(1/h_{it})\right) > 0$ and $\bar{y}_i^w \equiv w_i\left(\sum_{t=1}^{T} y_{it}/h_{it}\right)$, and a similar definition holds for $\bar{\mathbf{x}}_i^w$. Note that $\bar{y}_i^w$ and $\bar{\mathbf{x}}_i^w$ are simply weighted averages. If h_{it} equals the same constant for all t, $\bar{y}_i^w$ and $\bar{\mathbf{x}}_i^w$ are the usual time averages.

Now we can plug each $\hat{a}_i(\mathbf{b})$ into the SSR to get the problem solved by $\hat{\boldsymbol{\beta}}$:

$$\min_{\mathbf{b}\in\mathbb{R}^K} \sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - \bar{y}_i^w) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i^w)\mathbf{b}]^2/h_{it}.$$

But this is just a pooled weighted least squares regression of $(y_{it} - \bar{y}_i^w)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i^w)$ with weights $1/h_{it}$. Equivalently, define $\tilde{y}_{it} \equiv (y_{it} - \bar{y}_i^w)/\sqrt{h_{it}}$, $\tilde{\mathbf{x}}_{it} \equiv (\mathbf{x}_{it} - \bar{\mathbf{x}}_i^w)/\sqrt{h_{it}}$, all $t = 1,\ldots,T$, $i = 1,\ldots,N$. Then $\hat{\boldsymbol{\beta}}$ can be expressed in usual pooled OLS form:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}'\tilde{y}_{it}\right). \tag{10.82}$$

Note carefully how the initial y_{it} are weighted by $1/h_{it}$ to obtain $\bar{y}_i^w$, but where the usual $1/\sqrt{h_{it}}$ weighting shows up in the sum of squared residuals on the time-demeaned data (where the demeaming is a weighted average). Given (10.82), we can study the asymptotic ($N \to \infty$) properties of $\hat{\boldsymbol{\beta}}$. First, it is easy to show that $\bar{y}_i^w = \bar{\mathbf{x}}_i^w\boldsymbol{\beta} + c_i + \bar{u}_i^w$, where $\bar{u}_i^w \equiv w_i\left(\sum_{t=1}^{T} u_{it}/h_{it}\right)$. Subtracting this equation from $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$ for all t gives $\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}\boldsymbol{\beta} + \tilde{u}_{it}$, where $\tilde{u}_{it} \equiv (u_{it} - \bar{u}_i^w)/\sqrt{h_{it}}$. When we plug this in for $\tilde{y}_{it}$ in (10.82) and divide by N in the appropriate places we get

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}\tilde{\mathbf{x}}_{it}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}\tilde{u}_{it}\right).$$

Straightforward algebra shows that $\sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}\tilde{u}_{it} = \sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}u_{it}/\sqrt{h_{it}}$, $i = 1,\ldots,N$, and so we have the convenient expression

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}\tilde{\mathbf{x}}_{it}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}u_{it}/\sqrt{h_{it}}\right). \tag{10.83}$$

From (10.83) we can immediately read off the consistency of $\hat{\boldsymbol{\beta}}$. Why? We assumed that $\mathrm{E}(u_{it}|\mathbf{x}_i,\mathbf{h}_i,c_i) = 0$, which means u_{it} is uncorrelated with any function of $(\mathbf{x}_i,\mathbf{h}_i)$, including $\tilde{\mathbf{x}}_{it}$. So $\mathrm{E}(\tilde{\mathbf{x}}'_{it}u_{it}) = \mathbf{0}$, $t = 1,\ldots,T$. As long as we assume rank $\left(\sum_{t=1}^{T}\mathrm{E}(\tilde{\mathbf{x}}'_{it}\tilde{\mathbf{x}}_{it})\right) = K$, we can use the usual proof to show $\mathrm{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. (We can even show that $\mathrm{E}(\hat{\boldsymbol{\beta}}|\mathbf{X},\mathbf{H}) = \boldsymbol{\beta}$.)

It is also clear from (10.83) that $\hat{\boldsymbol{\beta}}$ is $\sqrt{N}$-asymptotically normal under mild assumptions. The asymptotic variance is generally

$$\mathrm{Avar}\ \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1},$$

where

$$\mathbf{A} \equiv \sum_{t=1}^{T}\mathrm{E}(\tilde{\mathbf{x}}'_{it}\tilde{\mathbf{x}}_{it}) \text{ and } \mathbf{B} \equiv \mathrm{Var}\left(\sum_{t=1}^{T}\tilde{\mathbf{x}}'_{it}u_{it}/\sqrt{h_{it}}\right).$$

If we assume that $\mathrm{Cov}(u_{it},u_{is}|\mathbf{x}_i,\mathbf{h}_i,c_i) = 0$, $t \neq s$, in addition to the variance assumption $\mathrm{Var}(u_{it}|\mathbf{x}_i,\mathbf{h}_i,c_i) = \sigma_u^2 h_{it}$, then it is easily shown that $\mathbf{B} = \sigma_u^2\mathbf{A}$, and so $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sigma_u^2\mathbf{A}^{-1}$.

The same subtleties that arise in estimating σ_u^2 for the usual fixed effects estimator crop up here as well. Assume the zero conditional covariance assumption and correct variance specification in the previous paragraph. Then, note that the residuals from the pooled OLS regression

$$\tilde{y}_{it} \text{ on } \tilde{\mathbf{x}}_{it},\ t = 1,\ldots,T,\ i = 1,\ldots,N, \tag{10.84}$$

say $\hat{r}_{it}$, are estimating $\tilde{u}_{it} = (u_{it} - \bar{u}_i^w)/\sqrt{h_{it}}$ (in the sense that we obtain $\hat{r}_{it}$ from $\tilde{u}_{it}$ by replacing $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$). Now $\mathrm{E}(\tilde{u}_{it}^2) = \mathrm{E}[(u_{it}^2/h_{it})] - 2\mathrm{E}[(u_{it}\bar{u}_i^w)/h_{it}] + \mathrm{E}[(\bar{u}_i^w)^2/h_{it}] = \sigma_u^2 - 2\sigma_u^2\mathrm{E}[(w_i/h_{it})] + \sigma_u^2\mathrm{E}[(w_i/h_{it})]$, where the law of iterated expectations is applied several times, and $\mathrm{E}[(\bar{u}_i^w)^2|\mathbf{x}_i,\mathbf{h}_i] = \sigma_u^2 w_i$ has been used. Therefore, $\mathrm{E}(\tilde{u}_{it}^2) = \sigma_u^2[1 - \mathrm{E}(w_i/h_{it})]$, $t = 1,\ldots,T$, and so

$$\sum_{t=1}^{T} \mathrm{E}(\tilde{u}_{it}^2) = \sigma_u^2\{T - \mathrm{E}[w_i \cdot \textstyle\sum_{t=1}^{T}(1/h_{it})]\} = \sigma_u^2(T - 1).$$

This contains the usual result for the within transformation as a special case. A consistent estimator of σ_u^2 is SSR/$[N(T - 1) - K]$, where SSR is the usual sum of squared residuals from (10.84), and the subtraction of K is optional. The estimator of Avar($\hat{\boldsymbol{\beta}}$) is then

$$\hat{\sigma}_u^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it}\right)^{-1}.$$

If we want to allow serial correlation in the $\{u_{it}\}$, or allow $\mathrm{Var}(u_{it}|\mathbf{x}_i,\mathbf{h}_i,c_i) \neq \sigma_u^2 h_{it}$, then we can just apply the robust formula for the pooled OLS regression (10.84).

CHAPTER 11

11.1. a. It is important to remember that, any time we put a variable in a regression model (whether we are using cross section or panel data), we are controlling for the effects of that variable on the dependent variable. The whole point of regression analysis is that it allows the explanatory variables to be correlated while estimating ceteris paribus effects. Thus, the inclusion of $y_{i,t-1}$ in the equation allows $prog_{it}$ to be correlated with $y_{i,t-1}$, and also recognizes that, due to inertia, y_{it} is often strongly related to $y_{i,t-1}$.

An assumption that implies pooled OLS is consistent is

$$\mathrm{E}(u_{it}|\mathbf{z}_i,\mathbf{x}_{it},y_{i,t-1},prog_{it}) = 0, \text{ all } t,$$

which is implied by but is weaker than dynamic completeness. Without additional assumptions, the pooled OLS standard errors and test statistics need to be adjusted for heteroskedasticity and serial correlation (although the later will not be present under dynamic completeness).

b. As we discussed in Section 7.8.2, this statement is incorrect. Provided our interest is in $\mathrm{E}(y_{it}|\mathbf{z}_i,\mathbf{x}_{it},y_{i,t-1},prog_{it})$, we do not care about serial correlation in the implied errors, nor does serial correlation cause inconsistency in the OLS estimators.

c. Such a model is the standard unobserved effects model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \delta_1 prog_{it} + c_i + u_{it}, \quad t=1,2,\ldots,T.$$

We would probably assume that $(\mathbf{x}_{it},prog_{it})$ is strictly exogenous; the weakest form of strict exogeneity is that $(\mathbf{x}_{it},prog_{it})$ is uncorrelated with u_{is} for all t and s. Then we could estimate the equation by fixed effects or first differencing. If the u_{it} are serially uncorrelated, FE is preferred. We could also do a GLS analysis after the fixed effects or first-differencing transformations, but we should have a large N.

d. A model that incorporates features from parts a and c is

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \delta_1 prog_{it} + \rho_1 y_{i,t-1} + c_i + u_{it}, \quad t = 1,\ldots,T.$$

Now, program participation can depend on unobserved city heterogeneity as well as on lagged y_{it} (we assume that y_{i0} is observed). Fixed effects and first-differencing are both inconsistent as $N \rightarrow \infty$ with fixed T.

Assuming that $\mathrm{E}(u_{it}|\mathbf{x}_i,\mathbf{prog}_i,y_{i,t-1},y_{i,t-2},\ldots,y_{i0}) = 0$, a consistent procedure is obtained by first differencing, to get

$$y_{it} = \Delta\mathbf{x}_{it}\boldsymbol{\beta} + \delta_1 \Delta prog_{it} + \rho_1 \Delta y_{i,t-1} + \Delta u_{it}, \quad t=2,\ldots,T.$$

At time t and $\Delta\mathbf{x}_{it}$, $\Delta prog_{it}$ can be used as there own instruments, along with $y_{i,t-j}$ for $j \geq 2$. Either pooled 2SLS or a GMM procedure can be used. Under strict exogeneity, past and future values of $\mathbf{x}_{it}$ can also be used as instruments.

11.3. Writing $y_{it} = \beta x_{it} + c_i + u_{it} - \beta r_{it}$, the fixed effects estimator $\hat{\beta}_{FE}$ can be written as

$$\beta + \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)\right)^{2}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(u_{it} - \bar{u}_i - \beta(r_{it} - \bar{r}_i)\right).$$

Now, $x_{it} - \bar{x}_i = (x_{it}^* - \bar{x}_i^*) + (r_{it} - \bar{r}_i)$. Then, because $E(r_{it}|\mathbf{x}_i^*, c_i) = 0$ for all t, $(x_{it}^* - \bar{x}_i^*)$ and $(r_{it} - \bar{r}_i)$ are uncorrelated, and so

$$\mathrm{Var}(x_{it} - \bar{x}_i) = \mathrm{Var}(x_{it}^* - \bar{x}_i^*) + \mathrm{Var}(r_{it} - \bar{r}_i), \text{ all } t.$$

Similarly, under (11.30), $(x_{it} - \bar{x}_i)$ and $(u_{it} - \bar{u}_i)$ are uncorrelated for all t. Now $E[(x_{it} - \bar{x}_i)(r_{it} - \bar{r}_i)] = E[\{(x_{it}^* - \bar{x}_i^*) + (r_{it} - \bar{r}_i)\}(r_{it} - \bar{r}_i)] = \mathrm{Var}(r_{it} - \bar{r}_i)$. By the law of large numbers and the assumption of constant variances across t,

$$N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i) \xrightarrow{p} \sum_{t=1}^{T}\mathrm{Var}(x_{it} - \bar{x}_i) = T[\mathrm{Var}(x_{it}^* - \bar{x}_i^*) + \mathrm{Var}(r_{it} - \bar{r}_i)]$$

and

$$N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(u_{it} - \bar{u}_i - \beta(r_{it} - \bar{r}_i) \xrightarrow{p} -T\beta\mathrm{Var}(r_{it} - \bar{r}_i).$$

Therefore,

$$\begin{aligned}\mathrm{plim}\ \hat{\beta}_{FE} &= \beta - \beta\left(\frac{\mathrm{Var}(r_{it} - \bar{r}_i)}{[\mathrm{Var}(x_{it}^* - \bar{x}_i^*) + \mathrm{Var}(r_{it} - \bar{r}_i)]}\right) \\ &= \beta\left(1 - \frac{\mathrm{Var}(r_{it} - \bar{r}_i)}{[\mathrm{Var}(x_{it}^* - \bar{x}_i^*) + \mathrm{Var}(r_{it} - \bar{r}_i)]}\right).\end{aligned}$$

11.5. a. $E(\mathbf{v}_i|\mathbf{z}_i,\mathbf{x}_i) = \mathbf{Z}_i[E(\mathbf{a}_i|\mathbf{z}_i,\mathbf{x}_i) - \boldsymbol{\alpha}] + E(\mathbf{u}_i|\mathbf{z}_i,\mathbf{x}_i) = \mathbf{Z}_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}) + \mathbf{0} = \mathbf{0}$. Next, $\mathrm{Var}(\mathbf{v}_i|\mathbf{z}_i,\mathbf{x}_i) = \mathbf{Z}_i\mathrm{Var}(\mathbf{a}_i|\mathbf{z}_i,\mathbf{x}_i)\mathbf{Z}_i' + \mathrm{Var}(\mathbf{u}_i|\mathbf{z}_i,\mathbf{x}_i) + \mathrm{Cov}(\mathbf{a}_i,\mathbf{u}_i|\mathbf{z}_i,\mathbf{x}_i) + \mathrm{Cov}(\mathbf{u}_i,\mathbf{a}_i|\mathbf{z}_i,\mathbf{x}_i) = \mathbf{Z}_i\mathrm{Var}(\mathbf{a}_i|\mathbf{z}_i,\mathbf{x}_i)\mathbf{Z}_i' + \mathrm{Var}(\mathbf{u}_i|\mathbf{z}_i,\mathbf{x}_i)$ because $\mathbf{a}_i$ and $\mathbf{u}_i$ are uncorrelated, conditional on $(\mathbf{z}_i,\mathbf{x}_i)$, by FE.1′ and the usual iterated expectations argument. Therefore, $\mathrm{Var}(\mathbf{v}_i|\mathbf{z}_i,\mathbf{x}_i) = \mathbf{Z}_i\boldsymbol{\Lambda}\mathbf{Z}_i' + \sigma_u^2\mathbf{I}_T$ under the assumptions given, which shows that the conditional variance depends on $\mathbf{z}_i$. Unlike in the standard random effects model, there is conditional heteroskedasticity.

b. If we use the usual RE analysis, we are applying FGLS to the equation $\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\alpha} + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i$, where $\mathbf{v}_i = \mathbf{Z}_i(\mathbf{a}_i - \boldsymbol{\alpha}) + \mathbf{u}_i$. From part a, we know that $\mathrm{E}(\mathbf{v}_i|\mathbf{x}_i,\mathbf{z}_i) = \mathbf{0}$, and so the usual RE estimator is consistent (as $N \to \infty$ for fixed T) and $\sqrt{N}$-asymptotically normal, provided the rank condition, Assumption RE.2, holds. (Remember, a feasible GLS analysis with any $\hat{\boldsymbol{\Omega}}$ will be consistent provided $\hat{\boldsymbol{\Omega}}$ converges in probability to a nonsingular matrix as $N \to \infty$. It need not be the case that $\mathrm{Var}(\mathbf{v}_i|\mathbf{x}_i,\mathbf{z}_i) = \mathrm{plim}(\hat{\boldsymbol{\Omega}})$, or even that $\mathrm{Var}(\mathbf{v}_i) = \mathrm{plim}(\hat{\boldsymbol{\Omega}})$.

From part a, we know that $\mathrm{Var}(\mathbf{v}_i|\mathbf{x}_i,\mathbf{z}_i)$ depends on $\mathbf{z}_i$ unless we restrict almost all elements of $\boldsymbol{\Lambda}$ to be zero (all but those corresponding the the constant in $\mathbf{z}_{it}$). Therefore, the usual random effects inference -- that is, based on the usual RE variance matrix estimator -- will be invalid.

c. We can easily make the RE analysis fully robust to an arbitrary $\mathrm{Var}(\mathbf{v}_i|\mathbf{x}_i,\mathbf{z}_i)$, as in equation (7.49). Naturally, we expand the set of explanatory variables to $(\mathbf{z}_{it},\mathbf{x}_{it})$, and we estimate $\boldsymbol{\alpha}$ along with $\boldsymbol{\beta}$.

11.7. When $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}/T$ for all t, we can rearrange (11.60) to get

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\lambda} + v_{it},\ t = 1,2,\ldots,T.$$

Let $\hat{\boldsymbol{\beta}}$ (along with $\hat{\boldsymbol{\lambda}}$) denote the pooled OLS estimator from this equation. By standard results on partitioned regression [for example, Davidson and MacKinnon (1993, Section 1.4)], $\hat{\boldsymbol{\beta}}$ can be obtained by the following two-step procedure:

(i) Regress $\mathbf{x}_{it}$ on $\bar{\mathbf{x}}_i$ across all t and i, and save the $1 \times K$ vectors of residuals, say $\hat{\mathbf{r}}_{it}$, $t = 1,\ldots,T$, $i = 1,\ldots,N$.

(ii) Regress y_{it} on $\hat{\mathbf{r}}_{it}$ across all t and i. The OLS vector on $\hat{\mathbf{r}}_{it}$ is $\hat{\boldsymbol{\beta}}$. We want to show that $\hat{\boldsymbol{\beta}}$ is the FE estimator. Given that the FE estimator can be obtained by pooled OLS of y_{it} on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$, it suffices to show that $\hat{\mathbf{r}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$ for all t and i. But

$$\hat{\mathbf{r}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{x}}_i'\mathbf{x}_{it}\right)$$

and $\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{x}}_i'\mathbf{x}_{it} = \sum_{i=1}^{N}\bar{\mathbf{x}}_i'\sum_{t=1}^{T}\mathbf{x}_{it} = \sum_{i=1}^{N}T\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i = \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i$, and so $\hat{\mathbf{r}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i\mathbf{I}_K = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$. This completes the proof.

11.9. a. We can apply Problem 8.8.b, as we are applying pooled 2SLS to the time-demeaned equation: $\text{rank}\left(\sum_{t=1}^{T}\text{E}(\ddot{\mathbf{z}}_{it}'\ddot{\mathbf{x}}_{it})\right) = K$. This clearly fails if $\mathbf{x}_{it}$ contains any time-constant explanatory variables (across all i, as usual). The condition $\text{rank}\left(\sum_{t=1}^{T}\text{E}(\ddot{\mathbf{z}}_{it}'\ddot{\mathbf{z}}_{it})\right) = L$ is also needed, and this rules out time-constant instruments. But if the rank condition holds, we can always redefine $\mathbf{z}_{it}$ so that $\sum_{t=1}^{T}\text{E}(\ddot{\mathbf{z}}_{it}'\ddot{\mathbf{z}}_{it})$ has full rank.

b. We can apply the results on GMM estimation in Chapter 8. In particular, in equation (8.25), take $\mathbf{C} = \text{E}(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{X}}_i)$, $\mathbf{W} = [\text{E}(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{Z}}_i)]^{-1}$, and $\boldsymbol{\Lambda} = \text{E}(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{u}}_i\ddot{\mathbf{u}}_i'\ddot{\mathbf{Z}}_i)$. A key point is that $\ddot{\mathbf{Z}}_i'\ddot{\mathbf{u}}_i = (\mathbf{Q}_T\mathbf{Z}_i)'(\mathbf{Q}_T\mathbf{u}_i) = \mathbf{Z}_i'\mathbf{Q}_T\mathbf{u}_i = \ddot{\mathbf{Z}}_i'\mathbf{u}_i$, where $\mathbf{Q}_T$ is the T x T time-demeaning matrix defined in Chapter 10. Under (11.80), $\text{E}(\mathbf{u}_i\mathbf{u}_i'|\ddot{\mathbf{Z}}_i) = \sigma_u^2\mathbf{I}_T$ (by the usual iterated expectations argument), and so $\boldsymbol{\Lambda} = \text{E}(\ddot{\mathbf{Z}}_i'\mathbf{u}_i\mathbf{u}_i'\ddot{\mathbf{Z}}_i) = \sigma_u^2\text{E}(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{Z}}_i)$. If we plug these choices of $\mathbf{C}$, $\mathbf{W}$, and $\boldsymbol{\Lambda}$ into (8.25) and simplify, we obtain

$$\text{Avar}\,\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sigma_u^2\{\text{E}(\ddot{\mathbf{X}}_i'\ddot{\mathbf{Z}}_i)[\text{E}(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{Z}}_i)]^{-1}\text{E}(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{X}}_i)\}^{-1}.$$

c. The argument is very similar to the case of the fixed effects estimator. First, $\sum_{t=1}^{T}\text{E}(\ddot{u}_{it}^2) = (T - 1)\sigma_u^2$, just as before. If $\hat{u}_{it} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}$

are the pooled 2SLS residuals applied to the time-demeaned data, then $[N(T - 1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^{2}$ is a consistent estimator of σ_u^2. Typically, $N(T - 1)$ would be replaced by $N(T - 1) - K$ as a degrees of freedom adjustment.

d. From Problem 5.1 (which is purely algebraic, and so applies immediately to pooled 2SLS), the 2SLS estimator of all parameters in (11.81), including $\boldsymbol{\beta}$, can be obtained as follows: first run the regression $\mathbf{x}_{it}$ on $d1_i$, ..., dN_i, $\mathbf{z}_{it}$ across all t and i, and obtain the residuals, say $\hat{\mathbf{r}}_{it}$; second, obtain $\hat{c}_1$, ..., $\hat{c}_N$, $\hat{\boldsymbol{\beta}}$ from the pooled regression y_{it} on $d1_i$, ..., dN_i, $\mathbf{x}_{it}$, $\hat{\mathbf{r}}_{it}$. Now, by algebra of partial regression, $\hat{\boldsymbol{\beta}}$ and the coefficient on $\hat{\mathbf{r}}_{it}$, say $\hat{\boldsymbol{\delta}}$, from this last regression can be obtained by first partialling out the dummy variables, $d1_i$, ..., dN_i. As we know from Chapter 10, this partialling out is equivalent to time demeaning all variables. Therefore, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ can be obtained from the pooled regression $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$, $\hat{\mathbf{r}}_{it}$, where we use the fact that the time average of $\hat{\mathbf{r}}_{it}$ for each i is identically zero.

Now consider the 2SLS estimator of $\boldsymbol{\beta}$ from (11.79). This is equivalent to first regressing $\ddot{\mathbf{x}}_{it}$ on $\ddot{\mathbf{z}}_{it}$ and saving the residuals, say $\hat{\mathbf{s}}_{it}$, and then running the OLS regression $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$, $\hat{\mathbf{s}}_{it}$. But, again by partial regression and the fact that regressing on $d1_i$, ..., dN_i results in time demeaning, $\hat{\mathbf{s}}_{it} = \hat{\mathbf{r}}_{it}$ for all i and t. This proves that the 2SLS estimates of $\boldsymbol{\beta}$ from (11.79) and (11.81) are identical. (If some elements of $\mathbf{x}_{it}$ are included in $\mathbf{z}_{it}$, as would usually be the case, some entries in $\hat{\mathbf{r}}_{it}$ are identically zero for all t and i. But we can simply drop those without changing any other steps in the argument.)

e. First, by writing down the first order condition for the 2SLS estimates from (11.81) (with dn_i as their own instruments, and $\hat{\mathbf{x}}_{it}$ as the IVs for $\mathbf{x}_{it}$), it is easy to show that $\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the IV estimator from (11.81) (and also (11.79)). Therefore, the 2SLS residuals from (11.81) are computed as $y_{it} - (\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}) - \mathbf{x}_{it}\hat{\boldsymbol{\beta}} = (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\boldsymbol{\beta}} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}$, which are exactly the 2SLS residuals from (11.79). Because the N dummy variables are explicitly included in (11.81), the degrees of freedom in estimating σ_u^2 from part c are properly calculated.

f. The general, messy estimator in equation (8.27) should be used, where $\mathbf{X}$ and $\mathbf{Z}$ are replaced with $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$, $\hat{\mathbf{W}} = (\ddot{\mathbf{Z}}'\ddot{\mathbf{Z}}/N)^{-1}$, $\hat{\mathbf{u}}_i = \ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Lambda}} = \left(N^{-1}\sum_{i=1}^{N}\ddot{\mathbf{Z}}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\ddot{\mathbf{Z}}_i\right)$.

g. The 2SLS procedure is inconsistent as $N \to \infty$ with fixed T, as is any IV method that uses time-demeaning to eliminate the unobserved effect. This is because the time-demeaned IVs will generally be correlated with some elements of $\mathbf{u}_i$ (usually, all elements).

11.11. Differencing twice and using the resulting cross section is easily done in Stata. Alternatively, I can used fixed effects on the first differences:

```
. gen cclscrap = clscrap - clscrap[_n-1] if d89
(417 missing values generated)

. gen ccgrnt = cgrant - cgrant[_n-1] if d89
(314 missing values generated)

. gen ccgrnt_1 = cgrant_1 - cgrant_1[_n-1] if d89
(314 missing values generated)

. reg cclscrap ccgrnt ccgrnt_1

  Source |       SS       df       MS                  Number of obs =      54
---------+------------------------------               F(  2,    51) =    0.97
   Model |  .958448372     2  .479224186               Prob > F      =  0.3868
Residual |  25.2535328    51   .49516731               R-squared     =  0.0366
---------+------------------------------               Adj R-squared = -0.0012
   Total |  26.2119812    53  .494565682               Root MSE      =  .70368

------------------------------------------------------------------------------
cclscrap |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
```

```
---------+--------------------------------------------------------------------
  ccgrnt |   .1564748   .2632934      0.594   0.555      -.3721087    .6850584
ccgrnt_1 |   .6099015   .6343411      0.961   0.341      -.6635913    1.883394
   _cons |  -.2377384   .1407363     -1.689   0.097      -.5202783    .0448014
------------------------------------------------------------------------------


. xtreg clscrap d89 cgrant cgrant_1, fe

                                                Fixed-effects (within) regression
sd(u_fcode)                  =   .509567                  Number of obs =       108
sd(e_fcode_t)                =  .4975778                              n =        54
sd(e_fcode_t + u_fcode)      =  .7122094                              T =         2

corr(u_fcode, Xb)            =   -0.4011                  R-sq within   =  0.0577
                                                               between  =  0.0476
                                                               overall  =  0.0050

                                                          F(  3,    51) =     1.04
                                                               Prob > F =  0.3827

------------------------------------------------------------------------------
 clscrap |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     d89 |  -.2377385   .1407362     -1.689   0.097      -.5202783    .0448014
  cgrant |   .1564748   .2632934      0.594   0.555      -.3721087    .6850584
cgrant_1 |   .6099016   .6343411      0.961   0.341      -.6635913    1.883394
   _cons |  -.2240491    .114748     -1.953   0.056      -.4544153    .0063171
------------------------------------------------------------------------------
   fcode |           F(53,51) =      1.674   0.033          (54 categories)
```

The estimates from the random growth model are pretty bad -- the estimates on the grant variables are of the wrong sign -- and they are very imprecise. The joint F test for the 53 different intercepts is significant at the 5% level, so it is hard to know what to make of this. It does cast doubt on the standard unobserved effects model without a random growth term.

11.13. To be added.

11.15. To be added.

11.17. To obtain (11.55), we use (11.54) and the representation $\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) = \mathbf{A}^{-1}(N^{-1/2}\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\mathbf{u}_i) + o_p(1)$. Simple algebra and standard properties of $O_p(1)$ and $o_p(1)$ give

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= N^{-1/2}\sum_{i=1}^{N}[(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \boldsymbol{\alpha}] \\ &\quad - \left(N^{-1}\sum_{i=1}^{N}(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\mathbf{X}_i\right)\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) \\ &= N^{-1/2}\sum_{i=1}^{N}(\mathbf{s}_i - \boldsymbol{\alpha}) - \mathbf{C}\mathbf{A}^{-1}N^{-1/2}\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\mathbf{u}_i + o_p(1)\end{aligned}$$

where $\mathbf{C} \equiv E[(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\mathbf{X}_i]$ and $\mathbf{s}_i \equiv (\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$. By definition, $E(\mathbf{s}_i) = \boldsymbol{\alpha}$. By combining terms in the sum we have

$$\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = N^{-1/2}\sum_{i=1}^{N}[(\mathbf{s}_i - \boldsymbol{\alpha}) - \mathbf{C}\mathbf{A}^{-1}\ddot{\mathbf{X}}_i'\mathbf{u}_i] + o_p(1),$$

which implies by the central limit theorem and the asymptotic equivalence lemma that $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is asymptotically normal with zero mean and variance $E(\mathbf{r}_i\mathbf{r}_i')$, where $\mathbf{r}_i \equiv (\mathbf{s}_i - \boldsymbol{\alpha}) - \mathbf{C}\mathbf{A}^{-1}\ddot{\mathbf{X}}_i'\mathbf{u}_i$. If we replace $\boldsymbol{\alpha}$, $\mathbf{C}$, $\mathbf{A}$, and $\boldsymbol{\beta}$ with their consistent estimators, we get exactly (11.55), sincethe $\hat{\mathbf{u}}_i$ are the FE residuals.